A Neural Approach to Cognition Based on Dynamic Field Theory

Jonas Lins and Gregor Schöner

1 Introduction

Much theoretical work has been dedicated to studying neural field equations at an abstract, mathematical level, focusing on the dynamic properties of the solutions (this book provides review of many of the latest efforts in this direction). Much less attention has been directed at understanding how function, in particular, how cognition may emerge from the different dynamic regimes and solutions of neural fields. This has left a gap between the mathematical models of neural fields that capture neurophysiology and theoretical models of neural function. Where is cognition in all the neural dynamics? May universal principles be identified that capture how the neural substrate rises to the demands of cognition?

Addressing these questions requires a theory that seamlessly integrates schemes of neural representation, fundamentals of cognition, and neural field dynamics. To ground the theory in neurophysiology, we must examine how neural activity captures and represents specific features of the world. To identify which properties of neural mechanisms enable them to support cognitive processes, we need a clear definition of cognition. Finally, a mathematical formalization must be chosen that endows neural fields with appropriate dynamical properties. In addition, such a mathematical theory must account for data from behavioral experiments that are observable indices of underlying cognitive processes.

In this chapter, we first address how patterns of neural activity represent attributes of sensory stimuli and of motoric actions. This provides the foundation for an operational theory of cognition based on dynamic neural fields, that we sketch next. We

Jonas Lins
Institute for Neural Computation, Ruhr-Universität Bochum, 44780 Bochum, Germany, e-mail: Jonas.Lins@ini.rub.de

Gregor Schöner
Institute for Neural Computation, Ruhr-Universität Bochum, 44780 Bochum, Germany, e-mail: Gregor.Schoner@ini.rub.de

review core concepts of Dynamic Field Theory (DFT; [27]), a theoretical framework that implements elementary forms of cognition as process models in neural field architectures, explains behavioral data, and generates testable predictions. We then briefly address the critical features of cognition, and discuss how DFT accounts for these properties. We conclude the chapter by describing an exemplary DFT architecture that illustrates how the sketched principles may be applied to model higher-level cognitive function.

2 Grounding DFT in Neurophysiology

To get a sense for how neural fields may represent percepts, actions, or cognitive states, we start with single neurons and then move to populations of neurons within a given area of the brain.

Neural tuning is the classical concept that links the activity of a neuron, located somewhere within the neural networks of the brain, with the external conditions to which the organism is exposed, either through sensory stimulation or through an action initiated by the organism. Most neurons in the higher nervous system are active only while a stimulus or motor parameter (e.g., color, shape, or movement direction) is within a restricted range. Within this range, the discharge rate is often a non-monotonic function of the parameter. These functions, called tuning curves, are obtained by plotting discharge rate against the manipulated parameter. In many cases, tuning curves are Gaussian or alike (although more complex schemes exist), centered around a "preferred" value of the parameter. For instance, neurons in visual cortex might respond vigorously to a particular direction of visual motion, while the spike rate falls off gracefully when the direction deviates more and more from that direction.

Tuning is found in the brain for a wealth of parameters. Classical examples include tuning to the position of stimuli on sensory surfaces, such as the location of a visual stimulus on the retina or of a tactile stimulus on the skin (tuning curves for spatial position are equivalent to receptive field profiles; [20] [29]). In other cases, it is motor space, such as the target position of a saccade [22] or the direction of a hand movement [10], that determines the activation of neurons. Cells may also be tuned to non-spatial feature dimensions like orientation [16] or color [4]. As a more complex example, neurons in visual area V4 are tuned to the curvature of object boundaries at specific angular positions relative to the object center [25]. In general, neurons tend to be sensitive and tuned to more than one dimension at the same time.

It is apparent from these examples that neurons often signal information about specific aspects of the sensed environment or of behavioral events. For brain areas close to the sensory or motor surfaces, it is relatively straightforward to determine tuning curves, by recording from neurons in a number of different sample conditions. Schwartz, Kettner and Georgopoulos [28], for instance, recorded the activity of 568 motor cortical neurons of monkeys while these performed hand movements in different directions. In each trial, the monkeys reached from a central position

to one of eight possible targets that were distributed in three-dimensional space. In the great majority of the recorded cells, discharge rate depended on the direction of movement. The tuning curves were well described by cosine functions of the angle formed between the current movement direction and the cell's preferred movement direction.

Tuning in motor cortex is therefore broad, relatively uniform, and the preferred values are broadly distributed, covering the entire dimension. As a result, the tuning curves of different neurons strongly overlap [10]. This is common in most areas of the brain and suggests that sensory or motor parameters are represented by neural populations: For any specific value of a sensory or motor parameter, say, a particular reaching direction, a large ensemble of neurons is active. The activation pattern induced by any individual stimulus or motor condition is best characterized as a distribution of activation within a neural population. But does the entire distribution matter, or do only the neurons contribute whose preferred values are closest to current parameter value, the neurons at the very center of the distribution?

According to the population coding hypothesis, information about currently coded parameter values is indeed represented jointly by all active neurons, with each neuron contributing according to its level of activation. Georgopoulos, Kettner, and Schwartz [11] tested this hypothesis for the coding of movement direction in motor cortex. Using the same experimental data as Schwartz et al. ([28]; see above), they determined a population vector [9] for each reaching movement and compared it to the actual movement direction. The population vector is a weighted vector sum of the preferred direction vectors of all active neurons, each preferred direction vector being weighted with the neuron's current spike rate (this is the theoretical mean of the distribution of population activation in circular statistics). The population vector turned out to be an excellent predictor of movement direction. Importantly, including more neurons in the population vector yielded more precise predictions, suggesting that indeed all neurons contributed to the behavioral outcome of the activation pattern.

Other findings lend more direct support to the population coding hypothesis by demonstrating that weakly activated neurons contribute to the coded estimate. In the superior colliculus, for instance, saccade targets are coded in a topographic map of retinal space. Saccades are rapid eye movements that serve to fixate a target position. Each saccade is accompanied by a blob of activity within the neural map, the position of which specifies the retinal target. Crucially, saccadic endpoints can be influenced by pharmacologically deactivating peripheral regions of the activity blob, even though neurons in these regions are only weakly activated, compared to the cells in the blob center [22]. Similarly, stimulation experiments in the middle temporal visual area (MT) show that the perception of visual motion direction is readily influenced by artificially induced activity peaks in the neural map of movement direction, even when the artificial peak is far from the visually induced peak [12]. Activity seems to be integrated across the whole map.

Apart from corroborating the population coding hypothesis, the population vector method is a first step toward an interpretation of neural population activity. However, the population vector reduces the entire distribution of activity to one single

value, discarding potentially meaningful information about its exact shape. Behaviorally relevant information potentially contained in multiple peaks of the distribution or in the shape of activation peaks is lost. Detecting the impact of these properties requires appropriate experimental paradigms and a method for constructing activity distributions from the firing of discrete neurons.

A way to do this is to construct the distribution of population activation (DPA; [5]). Although different variants of this method have been used [2] [17] [3], the basic rationale is to compute distributions from entire tuning curves rather than from the discrete preferred values of the neurons (Fig. 1).

First, the tuning curves of the recorded neurons are determined from a set of reference conditions (e.g., a sample of movement directions) and their amplitudes are normalized. The DPA of any particular test condition is then obtained as a weighted sum of these tuning curves. The weighting factors are the neurons' average spike rates in the test condition. The sum is normalized by the number of neurons and additional normalization steps compensate for uneven sampling from the distribution of preferred values. The DPA obtained this way is defined over the same parameter dimension as the tuning curves.

Bastian, Schöner, and Riehle [2] demonstrated that the shape of a DPA correlated with behavioral constraints. In the behavioral paradigm, monkeys reached from a central button to one of six target lights arranged around it in a hexagonal shape. Each trial started with the monkey pressing and holding the center button. A preparatory period followed, in which varied amounts of information about the upcoming movement direction were provided, by turning on one, two, or three adjacent lights. After one second one of the cued lights was turned to a different color as a definite response signal, prompting the monkey to move its hand to that light.
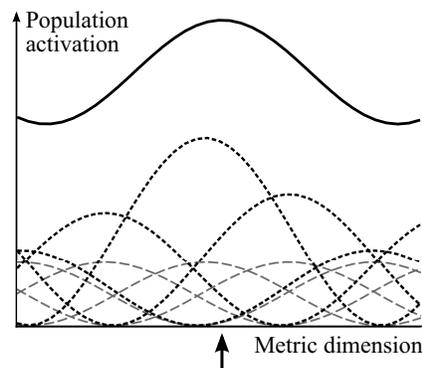
Fig. 1 To construct the distribution of population activation (DPA) for a particular parameter value (arrow), the tuning curves of the individual neurons (long dashed lines) are weighted by the respective neurons' firing rates. The weighted curves (short dashed lines) are then summed to obtain the DPA (solid line) over the coded dimension. Since neurons with preferred values close to the specified parameter value have higher firing rates, their curves contribute more strongly to the DPA. This leads to a peak at the position corresponding to the specified value, indicating that this value is currently represented by the population.

Immediately after the onset of the preparatory signal, a peak developed in the DPA, centered over the precued directions (Fig. 2). This activation remained above baseline throughout the preparatory period and increased slowly after the response signal was supplied. Thus, information about the potential movement directions was retained throughout this period. Moreover, the shape of the peak reflected the precision of prior information: more informative precues (i.e., fewer cue lights) led to higher and sharper peaks. When the definite response signal occurred, the peak sharpened and shifted toward the position corresponding to the cued movement direction, reaching its maximum height about 100 milliseconds before movement initiation. In this final stage, the shape of the peak was approximately equal for all conditions. In addition, Bastian and colleagues found that the shape of the DPA predicted reaction times (measured as the time from the occurrence of the response signal to movement onset): Broader, less pronounced peaks during the preparatory period corresponded to longer reaction times than sharp, pronounced peaks. Apparently, it took more time for broader peaks to reach sufficient concentration and height to initiate motor action. These findings show that the shape of the DPA impacts on behavior and that DPA shape may reflect different degrees of certainty or precision.

Using very similar techniques, Cisek and Kalaska [3] showed that multimodal DPAs may express different discrete choices of parameter values. They found that DPAs in premotor cortex can simultaneously represent two precued movement directions (only one of which is later realized). Clearly, the population vector is unable to represent such multi-valued information.
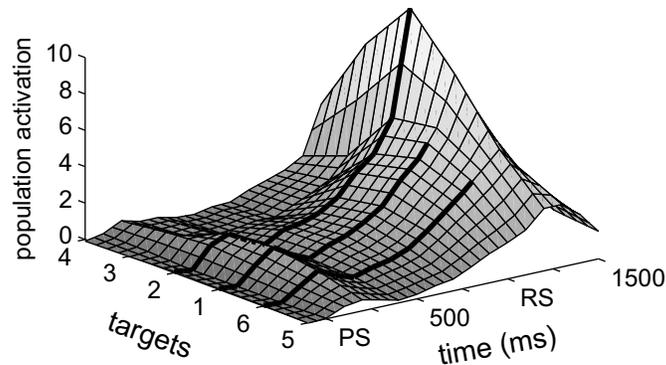
Fig. 2 Temporal evolution of a DPA during a reaching task, constructed from the firing of about 100 motor cortical neurons of a monkey [2]. The DPA is defined over the space of possible movement directions (targets). The DPA shown was obtained in a trial where the directional precue consisted of three adjacent lights (i.e., high directional uncertainty). The position of the three lights is indicated in the plot by thick black lines. From the occurrence of the preparatory signal (PS) onwards, a broad peak of activation emerges that is centered over the precued directions. It remains above activity baseline throughout the preparatory period. As the response signal (RS) occurs, activity sharpens, increases and shifts position, resulting in a pronounced peak centered around the final movement direction. Adapted from [2].

To summarize, peaks in DPAs pertain to macroscopically relevant perceptual or behavioral conditions, and the exact shape of the distribution carries information that may observably impact behavior. Thus, DPAs provide an appropriate level of consideration to assess the functional relevance of neural activity patterns.

3 Dynamic Field Theory

Dynamic Field Theory (DFT) builds on the finding that the relevant information is carried by distributions of activation among populations of neurons rather than by single cells. Via the DPA method, DFT is tightly linked to the physiology of population coding. DFT describes the evolution in time of activation patterns in neural populations. The activation patterns are modeled as dynamic neural fields (DNFs) that are defined over continuous metric dimensions and evolve continuously in time. The fields may be defined over virtually any perceptual, behavioral or cognitive dimension, such as color, retinal position, tone pitch, movement direction, or allocentric spatial position. Special focus is laid on modeling lateral neural interactions within the fields, endowing them with a particular set of stable attractor states. These stable states correspond to meaningful representational conditions, such as the presence or absence of a particular value along the coded dimension. Instabilities that lead to switches between the different stable states are brought about by sufficient changes in the configuration of the external input a field receives.

The particular mathematical form of field dynamics adopted by DFT has first been analyzed by Amari ([1]; see also [13] [34]):

$$\tau \dot{u}(x,t) = -u(x,t) + s(x,t) + h + \int w(x - x')\sigma(u(x',t)) \, dx' \qquad (1)$$

Here, $u(x,t)$ is the field of activation, defined over the metric dimension, x, and time, t. From a neurophysiological viewpoint, the activation, u, can be interpreted as a correlate to the mean membrane potential of a group of neurons. The time scale of the relaxation process is determined by τ. The field has a constant resting level, h, and may receive localized patterns of external input, $s(x,t)$. The last term describes lateral interactions between different field sites. Here, σ is a sigmoidal function implementing a soft threshold for field output, and w is an interaction kernel that specifies the strength of interactions between different field sites as a function of their metric distance. The kernel typically has a Mexican hat shape, implementing local excitation and surround inhibition, usually with added global inhibition. This means that field sites coding for similar parameter values excite each other, while mutual inhibition predominates between field sites that code for very different values. The sigmoidal threshold function ensures that only sufficiently activated field sites generate output and impact on other sites. The field output can be viewed as corresponding to the mean spike rate of a group of neurons.

In the absence of supra-threshold activation, no output is generated. In this case, the entire field relaxes to the stable attractor that is set by the resting level (which

usually resides well below the output threshold). A flat distribution indicates the absence of any specific information about the coded dimension.

When weak, localized input is applied, the attractor at the respective field site is shifted toward the output threshold. As long as the threshold is not reached, though, the field state remains purely input-driven and activation thus simply traces the shape of the input (Fig. 3A). Although there is now some structure to the distribution, this state still indicates the absence of conclusive information.

If, in contrast, the localized input is sufficiently strong to push a section of the field above threshold, output is generated and lateral interaction kicks in. Provided the parameters of the interaction kernel are within an appropriate range, lateral interaction promotes the formation of a localized peak of activation (Fig. 3B). Local excitation further elevates activation around the input position, whereas more distant field sites are depressed by global inhibition and/or surround inhibition, which prevents the peak from dispersing. Due to these properties we refer to this as a self-stabilized peak.

We call the transition from a sub-threshold solution to a self-stabilized peak the detection instability, because it corresponds to the decision that a coherent, well-defined item is present in the input stream. Peaks are units of representation in this sense, indicating that a particular parameter value is present in the sensory environment, as part of a motor plan, or as the contents of memory. The encoded value itself – what is being perceived, planned, or memorized – is specified by the position of the peak along the metric dimension. Peak height and width, on the other hand, may reflect certainty, intensity and precision (analogous to DPAs).

Conversely, we call it the reverse detection instability when an existing self-stabilized peak vanishes. This happens when the localized input that brought about the peak is sufficiently reduced in strength. For example, when the input is removed entirely, the peak attractor becomes unstable and disappears, while the resting level attractor reappears, to which the system then relaxes. Decreasing the input strength successively will also eventually trigger the reverse detection instability. However, local excitation to a degree shields existing peaks from decaying. The system will thus stick to the detection decision across a range of input strengths that would not have triggered the detection instability in the first place. The system is bistable over this range, with the peak attractor and the input-driven attractor coexisting (Fig. 3C). The field state then depends on which basin of attraction it resided in prior to the change of input strength. In other words, the state of the system depends on its activation history, it displays hysteresis.

Hysteresis stabilizes decisions against random fluctuations and perturbations. In the nervous system, such fluctuations may arise due to the inherent variability in neural firing or as the result of currently ongoing but unrelated neural processes. Under such conditions a lack of hysteresis would lead to constantly fluctuating decisions for near-threshold input. It is thus unsurprising that signatures of hysteresis are a common finding in behavioral experiments (for review, see [15]). For instance, perceptual hysteresis has been reported for single-element apparent-motion [14]. In the experiment, participants were shown two squares with differing luminance. While the participants were watching, the squares constantly exchanged their lumi-

nance values, which created a percept of either flicker or apparent motion between them. In each trial participants reported at multiple time points which of the two percepts they currently experienced. The decisive variable predicting whether motion or flicker was perceived was background-relative-luminance contrast (BRLC).
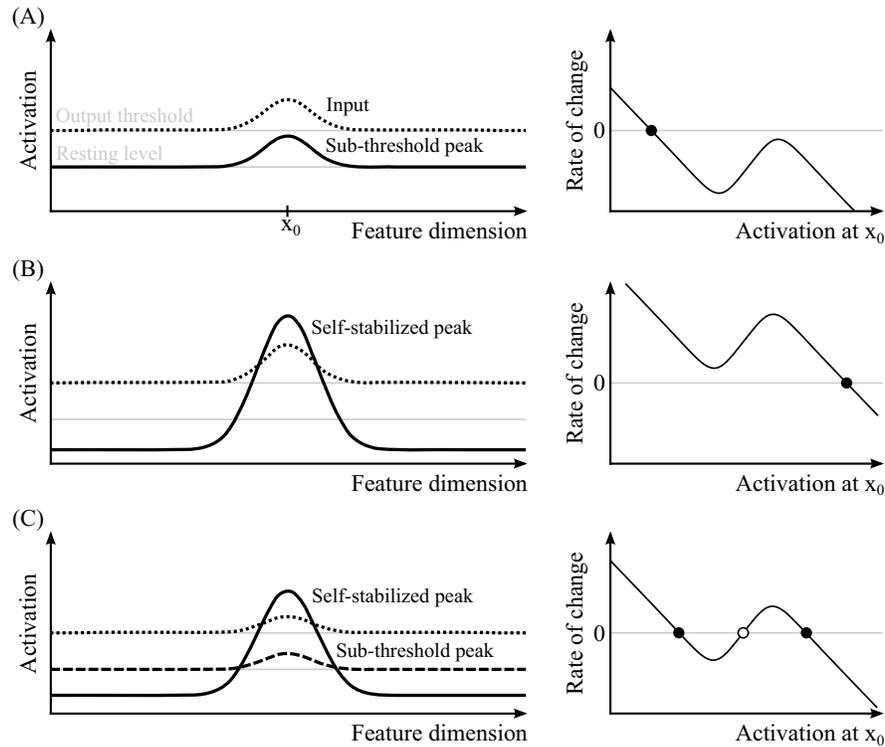
Fig. 3 Left column: Stable states reached by dynamic neural fields (solid lines and long dashed line) as a result of localized Gaussian inputs of different strengths (dotted lines). Right column: Corresponding plots of the rate of change as a function of activation at the peak position, x_0 (note that these plots are only approximate, as they do not take into account the impact of other field sites on the rate of change at x_0 via lateral interactions). Attractors are marked by filled dots, repellors by open dots. (A) Weak input results in a purely input-driven sub-threshold peak, which is a monostable attractor state. (B) High levels of input that bring activation above threshold result in output generation and lateral interactions, thus leading to a self-stabilized peak. This state as well is monostable. (C) For intermediate input strengths the system reaches a bistable state. The current state then depends on the system's prior state. Here, the self-stabilized peak (solid line) corresponds to the attractor on the right side, which is reached from high levels of activation. The sub-threshold peak (long dashed line) corresponds to the left attractor, which is reached from low levels of activation.

BRLC is the strength of luminance change between individual frames in relation to how much the spots' average luminance differs from background luminance. High BRLC led to motion percepts more frequently than low BRLC. Hysteresis was observed when BRLC was changed continuously during the trials in a descending or ascending manner. The BRLC value at which the motion percept was lost in descending trials was lower than the value at which the motion percept was established in ascending trials. This suggests that within a certain range of BRLC values motion perception is bistable.

Another fundamental attractor state in DFT is the self-sustained peak state. Self-sustained peaks arise in much the same way as self-stabilized ones: When localized, excitatory input brings activation above threshold, output is generated, driving lateral interactions that support peak formation. The difference lies in the balance of excitation and inhibition. A field supports self-sustained peaks if local excitation is sufficiently strong, relative to inhibition, to by itself prevent peaks from decaying after the input is removed. In this regime, peaks decay only when the level of activation is sufficiently decreased, locally or globally, by external inhibitory input or by endogenous inhibitory interactions. Otherwise, self-sustained peaks may persist indefinitely, even in the absence of localized input. The self-sustained regime enables DNFs to support the functionality of the neural process of working memory (see also, [6], and the original [7]), which will be illustrated in Section 5.

We have so far considered only single localized inputs. But natural environments are usually richly structured – visual scenes are cluttered with objects, auditory signals arrive from multiple directions, and so forth – which in turn implies a variety of potential behavioral goals and movement targets. Several inputs may be equally salient due to, say, equal brightness, contrast, or loudness. Under such conditions multiple stimuli compete for processing and behavioral impact. In terms of DNFs, this amounts to a field receiving multiple localized inputs at the same time. DFT provides dynamic mechanisms that mediate selection in such situations.

The selection of saccade targets is a well-studied example, which has been addressed in detail by DFT modeling efforts. We base the following considerations on those efforts, mainly on a model by Wilimzig, Schneider, and Schöner [33], which in turn complements prior modeling work [21] [32].

Saccades are rapid eye movements that serve to quickly fixate targets in visual space. Saccade trajectories are planned prior to the initiation of the movement and are not adjusted afterwards. The metrics of saccades are specified in the superior colliculus, a mid-brain structure that integrates cortical and direct visual input in a topographic map of visual space. Activation bumps in this map specify the vertical and horizontal extent of saccades. It is thought that the superior colliculus plays an important role in both target selection and saccade initiation. The DNFs in the model sketched below can be viewed as roughly corresponding to the respective neural populations in the superior colliculus.

The model consists of a selection level and an initiation level. Each level is constituted by a DNF with local excitation and global inhibition, defined over the space of saccadic endpoints (i.e., retinal space). The space is modeled as one-dimensional, which is sufficient to capture most experimental paradigms. We will restrict our con-

siderations to the selection level, where selection between different visual targets occurs.

Items in the visual field are fed into the selection field as localized, Gaussian-shaped input patterns. When a single target item generates sufficient input, the detection instability occurs, resulting in the formation of a self-stabilized peak of activation in the field. Although the model of Wilimzig et al. accounts for several experimental findings with regard to the case of a single target as well, we are here primarily interested in situations in which at least two targets are presented simultaneously. What happens in the double target case depends on the exact configuration of the two inputs and on activation biases that may be caused by stochastic perturbations or by imbalances between the two stimuli.

We first consider the case of two visual items that are equally salient and spatially remote from each other. Also, we will assume that there are no stochastic perturbations, and thus no random activation biases. Due to their equal saliency, both items generate input of the exact same strength. Furthermore, the large distance between the items ensures that there is no interaction between the resulting peaks, except for homogenous global inhibition. As a consequence, two supra-threshold peaks emerge that are somewhat less pronounced than in the single input case, as the sum of global inhibition is larger. The resulting field state is a fixed point of the system, but it is not stable.

This becomes apparent when an activation bias is introduced. One source of such imbalance are stochastic perturbations caused by neuronal variability or by other ongoing neural processes (the saccade model implements stochastic variability as Gaussian white noise). Random fluctuations of activity may provoke selection decisions by strengthening or weakening one of the competing peaks. Another source of imbalance is the relative strength of the inputs themselves. In a visual context, the strength of an input may be associated, for instance, to the brightness or contrast of a stimulus. Regardless of its source, an imbalance in favor of one of the peaks leads to increased activation and excitation around this location, increasing the height of the peak above that of its competitor. The ensuing increase in global inhibition suppresses the weaker peak, eventually reducing it to an input-driven bump. The single-peak state resulting from this selection decision is bistable, with both peak attractors coexisting (Fig. 4A).

However, multiple inputs do not always result in selection, but may also lead to fusion. This has been shown empirically for the case of saccades by Ottes, van Gisbergen, and Eggermont [24]. Their participants made saccades from a fixation point to a green stimulus whenever it appeared. In some trials, a red stimulus appeared alongside the green stimulus. Although the participants were instructed to ignore the red item, the first saccade they made often landed at an average position between the two stimuli. Such averaging saccades occurred much more frequently when the stimuli were spatially close than when they were widely separated.

This phenomenon is as well captured by the saccade model. It is, in fact, a feature of DNFs in general. As observed by Ottes et al., whether selection or fusion occurs depends to a large part on the inputs' metrics. Specifically, two inputs tend to result in a single peak at an average position if they are so close to each other that the

regions of input-induced activation are subject to mutual excitation. In that case, the activation propagates from the two input positions towards the center between them, eventually forming a single peak (Fig. 4B). The state with fused peak is monostable for very close inputs. When the distance between them is increased after a fused peak has already established, it becomes bistable at some point. That is, although the fused peak persists, applying the same input configuration to a previously inactive field would result in a selection decision. If the distance is increased even more, the attractor of the fused peak eventually becomes unstable and disappears. The field then relaxes to the now monostable selection state. We call this the fusion/selection instability.

In the model of Wilimzig et al. inhibition is not implemented via the same interaction kernel as excitation, but it is instead mediated by an additional layer of interneurons. This layer receives excitatory input from the main, purely excitatory field and projects back global inhibition to it. On the one hand, this implementation of inhibition was chosen to accord to Dale's law, which states, roughly put, that each neuron releases the same set of neurotransmitters at all of its synapses. Assuming that DNFs capture homogeneous neural populations, this means that a field's output can be either excitatory or inhibitory, but not both. On the other hand, using a separate inhibitory field has the effect of delaying the impact of inhibition compared to that of local excitation. Due to the different roles of excitation and inhibition in
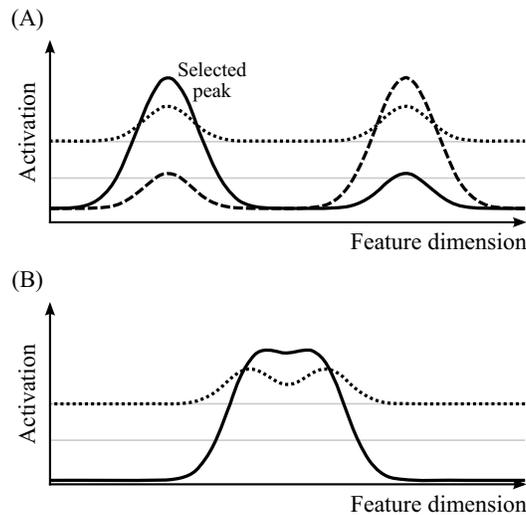
Fig. 4 Stable states reached by dynamic neural fields (solid lines and long dashed line) as a result of different patterns of localized Gaussian input (dotted lines). (A) Competition between peaks occurs when two inputs are applied at distant positions. Only at one location is a self-stabilized peak formed (solid line), while the other is suppressed by inhibition. The state resulting from this selection decision is bistable, with the alternative state (long dashed line) continuing to coexist as an attractor. Which state is reached depends on the field's prior activation history, imbalances between the inputs, and noise. (B) Two close inputs can result in a monostable fused peak state, with a single peak at an average location between the inputs.

the fusion and selection of peaks this gives rise to a specific association between the latency and the type of saccades. Saccades with lower latencies are more likely to target an average position between the stimuli, while later saccades tend to select one target. This effect has also been found empirically [24].

Note that the maximum number of peaks that a DNF can support depends on the balance of excitation and inhibition. Fields in DFT are not generally constrained to a single peak. However, the number of peaks is usually quite limited through inhibition. This is particularly relevant when modeling explicit capacity limits in cognition, such as those in working memory or attentional function.

4 DFT as an Approach to Cognition

Through the dynamic properties described in the previous section, DNFs acquire capabilities that are at the core of cognition: making decisions and maintaining the outcome of these decisions. Detection decisions make neural representations to some extent independent from the continuous input stream. Selection decisions further decouple the contents of neural representations from the immediate input, by separating items into those that impact processing and those that are ignored. The specific stability properties of DNFs ensure that the outcomes of these decisions are shielded from changes in the input, and retained as long as needed.

The paramount importance of these capabilities for cognition is perhaps best illustrated by considering an alternative approach, one that may seem more straightforward at first glance, but that faces profound issues when it comes to cognition – exactly because it lacks the capabilities described above.

To start with, the nervous system is immersed in a continuously changing environment, facing a continuous stream of sensory input. A simple way to guide behavior based on this type of input is to use some form of continuous closed-loop control system. That is, a mechanism that continuously maps sensory input and feedback to motor action, according to an appropriate function. Such systems may perform intriguingly complex control tasks. Many simple biological organisms work this way, as well as many systems of the human body, such as the regulation of blood pressure. Could cognition be nothing more than the input-driven evolution of activation in a system of this type? To put it more radically, may cognition work on the same principles as the regulation of blood pressure? Such a view has indeed been advanced by some (e.g., [8]).

The problem with such systems is that the linkage between their different subparts is seamless from input to output. The value of every variable is uniquely specified at any point in time and there are no discontinuities between the input and the output stage. As a consequence, subparts of such systems cannot shield their own state from the continuous input stream or from the impact of other subparts. So the output of the system is basically a transformed version of the input stream that is tightly coupled to the input at all times. There is also no way of how such a system may represent the absence of information. Control variables cannot be "empty".

Conversely, the emergence of new states is always strongly affected by previous states. Due to these properties, the emergence of decisive hallmarks of cognition and behavior cannot be explained by such systems.

One of these hallmarks is the discreteness of behavioral events. How may discrete behavioral events be initiated and terminated on the basis of purely continuous processes? There must be a gap somewhere in between continuous input, intermediate processes, and behavior, that cannot be explained in terms of systems as the one described above.

On a closer look, the problem does not apply to motor action alone, but extends to those capabilities of the brain that are often regarded as "higher" forms of cognition. For example, mental imagery, working memory, sequence generation – all these faculties have in common that their functioning requires a degree of independence from the current sensory or motor environment. Working memory initially requires sensory input to store, but after storage has been achieved, it requires that the stored information be shielded from being overwritten by new input. Imagery is essentially defined by the independence of a perceptual brain state from current sensory input. Actions in a controlled sequence that works toward some distal goal (e.g., making coffee) need to be shielded against distracting input that would trigger unrelated behavior (e.g., taking a cup and cleaning it).

Thus, at the heart of cognition lies the nervous system's capability to generate, maintain, and act upon inner states that are, to a degree, independent from current sensory input. Mechanisms are needed that decouple the representations upon which cognitive operations are carried out from the immediate sensed world (and from each other). On the other hand, behavior and cognition still need to be closely linked to the sensory surfaces, else we would think and act completely aloof. This is the core assumption of the stance of embodied cognition [26]. Cognition and behavior are still flexible, in that they can be updated online if relevant new input is detected. So what is needed is a balance between decoupling and coupling that allows only certain input to impact on downstream systems, cognition, and action, but which nonetheless allows these decisions to be changed if appropriate.

DFT effectively implements these demands. Note that the decisive capabilities, detection, selection, and appropriate stability, are realized in each individual field. This means that elementary forms of cognition happen at already very low levels of computation rather than being dependent on complex architectures. We have considered some concrete examples for this in the previous section (e.g., in the context of the saccade model). However, the stability of individual fields also allows for the construction of modular field architectures, which may implement more complex cognitive tasks and employ layers farther removed from the immediate sensory and motor surfaces. We conclude the chapter by considering such an architecture.

5 Modeling Visual Working Memory and Change Detection With Dynamic Neural Fields

Our considerations are based on a model originally proposed by Johnson and colleagues [18] [19] that is rooted in a general DFT approach to visual and spatial cognition [30] [31]. The model addresses the two closely linked cognitive domains of visual working memory and change detection. Visual working memory stores recent visual input over durations in the order of seconds and makes this information available to other processes. Change detection means comparing the contents of visual working memory to newly incoming visual input. Combining visual working memory and change detection yields a strategy for detecting changes in visual scenes despite the frequent interruptions of the visual input stream by saccades and blinks.

Change detection can be probed experimentally by showing to the participant a display with several simple visual items, such as colored dots, that differ along at least one feature dimension (e.g., [23]). After a short delay of normally less than a second a test display is presented that is either identical to the previous one or in which one of the items has changed with respect to one feature (e.g., an item may have changed color). Participants then indicate whether or not they perceive a change, by responding "different" or "same".

To perform this task, the items in the first display need to be perceived and encoded into working memory. The resulting representation must then be shielded from new input and maintained over the delay. Finally, the contents of visual working memory must be compared to the test display, which requires integrating working memory and perception.

As a first step in accomplishing this, a system must be capable of retaining information in the absence of input. Self-stabilized peaks are well-suited for perception, since they are quite tightly linked to the presence of input, but they decay when the input is removed for a longer period of time. Working memory therefore requires a different dynamic regime, namely, the self-sustained one. Because DNFs can operate in only one regime at a time, perception and working memory require separate fields. The model by Johnson and colleagues thus employs a perceptual field and a working memory field (Fig. 5).

The different functional roles of these fields arise from their different sources of excitatory input (dashed gray line and solid arrows in Fig. 5) and the different dynamic regimes in which they operate. The perceptual field receives direct feature input from the visual scene (taking the form of Gaussians) and operates in a self-stabilized regime. The working memory field receives its main excitatory input from the perceptual field (and weak direct visual input) and operates in a self-sustained regime. The integration of perception and working memory is achieved through a shared layer of inhibitory interneurons. This layer operates in a purely input driven regime and receives excitatory input from both the perceptual and the working memory field. In turn, the inhibitory layer sends back broad (but localized) inhibition to both other fields (dashed arrows in Fig. 5). That is, surround inhibition in both the

perceptual and the working memory field is mediated by the inhibitory interlayer, as well as mutual inhibition between the perceptual and the working memory field. All fields are defined over a metrically scaled visual feature dimension.

We first consider how a single feature input (corresponding to, say, a single colored item) is encoded perceptually, encoded to working memory, maintained over a delay, and compared to a subsequent input. For this we refer to the simulation results shown in Figure 6.

In model terms, presenting a visual item amounts to providing localized input to the perceptual field (and a much weaker version of the same input to the working memory field). If the input is sufficiently strong, the perceptual field undergoes the detection instability, leading to a self-stabilized peak (Fig. 6, t = 400). This step corresponds to the perceptual encoding of the input feature value. Once the peak has established, the perceptual field provides localized input to the working memory field, leading to a self-sustained working memory peak. This corresponds to encoding the feature value into working memory. As soon as the external input is removed, the perceptual peak destabilizes and decays, while the working memory peak is maintained over the delay in the absence of input (Fig. 6, t = 800).

The change detection functionality naturally emerges from this setup. Mediated by the inhibitory layer, the sustained working memory peak leads to inhibition of the perceptual field at the field site corresponding to the feature value held in memory. The resulting activation trough is critical, because it ensures that new input to the
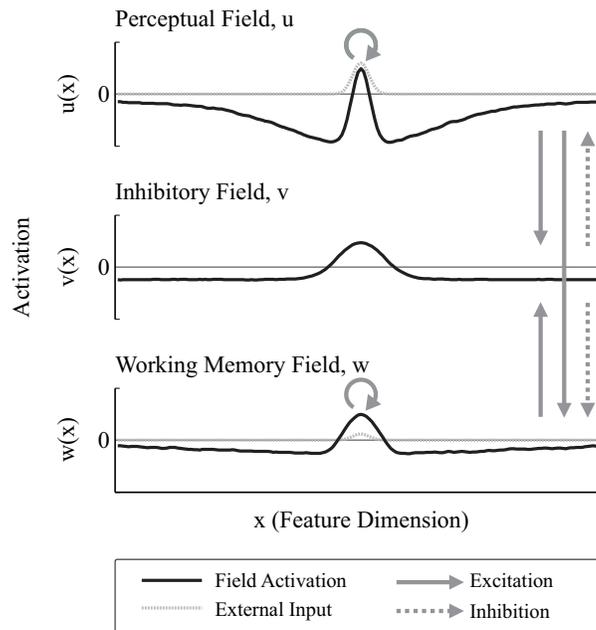
Fig. 5 The three-layer DNF architecture of visual working memory and change detection [19]. See text for details.
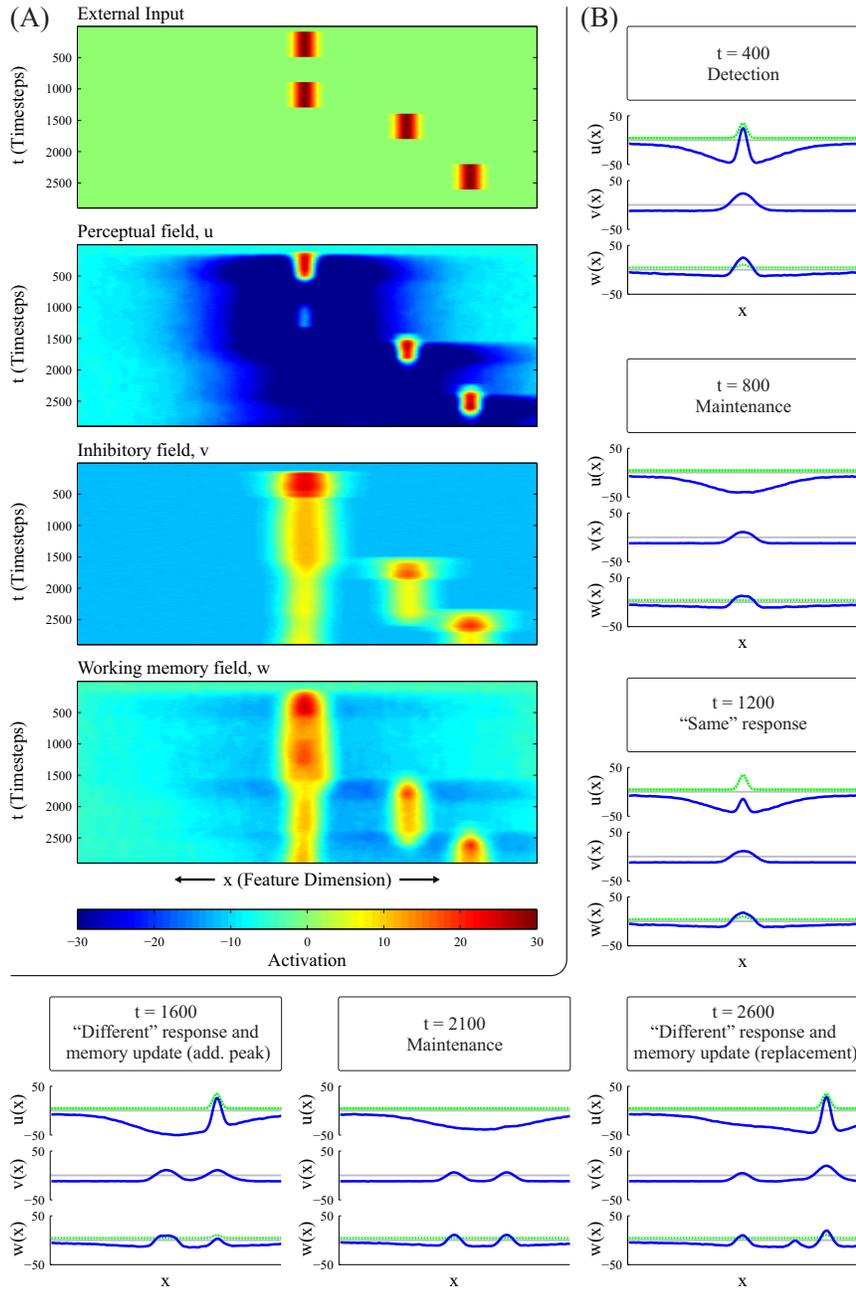
Fig. 6 Simulation of change detection in the three-layer DNF architecture. (A) Evolution of field activation over simulation time steps, in response to the pattern of feature input shown in the topmost plot. (B) Snapshots of the model state at selected time steps. At t = 400 the single localized input (dashed green line) has led to a self-stabilized peak in the perceptual field, which projects to both the working memory field and the inhibitory field. Due to the self-sustaining regime of the working memory field, the peak there is maintained in the absence of input (t = 800). Through the inhibitory field it creates a trough of inhibition in the perceptual field. If at test the same or a very similar item is shown (t = 1200) the input-driven hump in the perceptual field coincides with a region of the trough where there is strong inhibition. This makes it unlikely that the output threshold is reached. If, in contrast, the difference between the test item and the retained item is sufficiently large (t = 1600), the input impacts on a less strongly inhibited region of the perceptual field so that the threshold is reached more easily. The projection of the new perceptual peak to the working memory field may lead to the updating of working memory with an additional peak that is then as well maintained over periods without input (t = 2100). Alternatively, an existing memory peak may be suppressed and replaced if the new input is relatively close to an existing one (t = 2600).

perceptual field reaches threshold only if the test item is sufficiently different from the value held in memory. The change detection mechanism thus comes into effect in a completely autonomous manner when new input arrives. If the test item is very similar to the retained one, the visual input to the perceptual field coincides with the center of the trough of inhibition (Fig. 6, t = 1200). This makes it unlikely that the output threshold is reached. Accordingly, the absence of a peak in the perceptual field and the concurrent presence of a peak in the working memory field at test time indicate that no change has been detected ("same" response). If, in contrast, the test item is metrically sufficiently different from the first one, the visual input peak is somewhat displaced from the center of the trough (Fig. 6, t = 1600). It thus impacts on a field site where inhibition is less pronounced, so that the output threshold is reached more easily. Therefore, supra-threshold activation in the perceptual field at test time means that a change has been detected ("different" response).

Note that the same/different decision can be made explicit by introducing two self-excitatory, mutually inhibitory dynamical nodes, a "same" node and a "different" node. The "different" node receives summed activation from the perceptual field, while the "same" node receives summed activation from the working memory field. To force a decision at test time, a boost of activation is applied to both nodes, leading to the selection of one alternative, depending on the ratio of supra-threshold activation in the perceptual and the working memory field. These nodes were introduced to enable comparisons of the model performance with behavioral data [18]. For simplicity we omit this detail in our considerations.

Apart from signaling change, a peak in the perceptual field may have the effect of updating the contents of working memory. Depending on the metric distance between the already existing peaks and the new input, this can either mean that a completely new peak is established (for remote items; Fig. 6, t = 1600), leading to a stable multi-peak solution (Fig. 6, t = 2100), or that an existing peak destabilizes and is replaced by the new peak (for close items; Fig. 6, t = 2600).

The functionality we have described so far generalizes to multiple items. When several different inputs are applied in the encoding phase, a multi-peak solution arises and persists in the working memory field (Fig. 7A, B). Change detection then works in the same way as for single items. However, with a larger number of items, interactions between the retained peaks can lead to effects not observed in the case of a single item. For example, because the level of inhibition increases with the number of peaks in working memory, there is a capacity limit with respect to the total number of items that can be retained (about four in this particular model). This limit can lead to the deletion of existing peaks by new input ("forgetting"), to incomplete encoding of multi-item displays, or to failure to encode new input. The all-or-none property of working memory in the model – a stable peak is either formed or not – is consistent with behavioral data [35].

A counter-intuitive prediction made by the model is that change detection should be enhanced when two metrically close items are retained in working memory and a new, slightly different item is presented. This is because inhibition in the model is not global, but local, and tied to the position of peaks, so that nearby peaks inhibit

each other more strongly than other peaks. Close peaks in working memory are thus less pronounced than more isolated ones (Fig. 7B).

Because the peaks are smaller, they also lead to slightly less inhibition and a shallower trough in the perceptual field. This makes it more likely that new input within the area of the trough reaches threshold and generates a "different" response. Therefore, the same degree of difference between a test item and a retained item can result in either a "same" or a "different" response, depending on whether the nearest working memory peak is relatively isolated (Fig. 7C) or has other peaks in its vicinity (Fig. 7D). This prediction has been confirmed empirically [18].

Other behavioral evidence that has been successfully captured by the model includes the selection of inputs for encoding into working memory (e.g., based on saliency) and the mutual repulsion of retained values along the feature dimension.

To summarize, the model provides a neurally plausible process account for visual working memory and change detection, captures diverse behavioral data, and has been a source of new, testable predictions. This example illustrates that, by virtue of
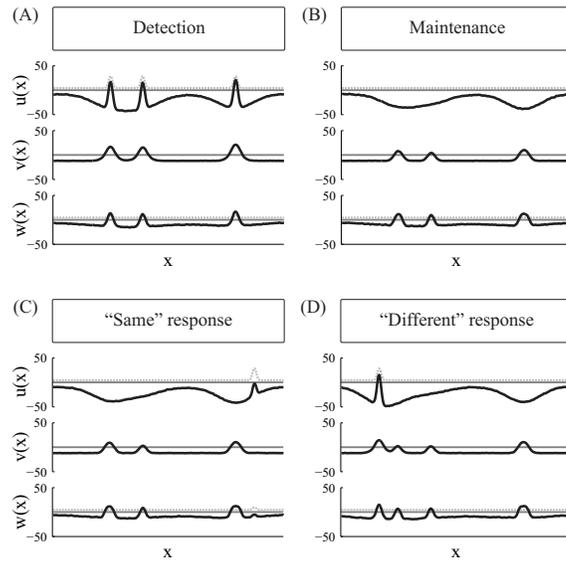
Fig. 7 Enhancement of change detection in the three-layer architecture due to similar items being held in working memory. (A) Encoding of three items into working memory, two close ones and an isolated one. (B) The two close peaks in working memory inhibit each other and are thus less pronounced than the isolated one. In turn, the trough of inhibition in the perceptual field caused by the isolated peak is slightly more pronounced than the trough caused by the joint impact of the two close peaks. (C) If the item shown at test is similar to the isolated peak, the input coincides with the deeper trough, making an erroneous "same" response more probable. (D) Here the test item is instead similar to one of the close items. Although the degree of similarity is the same as in (C), the threshold is reached more easily since the new input falls into a region of the shallower trough, having to overcome less inhibition.

its modularity, DFT is well suited to capture not only elementary aspects of cognition, but also more complex (or "higher") cognitive acts.

6 Conclusions

We have reviewed how neural fields may be viewed as mathematical descriptions of distributions of population activation. Their dynamics, captured in Dynamic Field Theory, leads to a set of stable states, sub-threshold solutions, self-stabilized peaks, self-sustained peaks, and associated instabilities, the detection, selection, and memory instability. From these, cognitive properties of dynamic neural processes emerge. By linking neural field dynamics to behavioral signatures of sensory, motor, and cognitive function, DFT provides an interface between neurally grounded process models and cognition. The instabilities of DFT provide critical properties of cognitive processes, most prominently, the capability to both isolate cognitive states from distractor input or interaction, while at the same time maintaining the capacity to link cognitive processes to ongoing sensory and motor processes as well as to other concurrent cognitive processes.

Much work remains to be done to ground all cognition in neural processing. DFT has helped make the first steps, emphasizing the embodied nature of cognition. The frontier now is to move such principles toward higher cognition.

References

1. Amari, S.: Dynamics of pattern formation in lateral-inhibition type neural fields. Biological cybernetics **27**(2), 77–87 (1977)
2. Bastian, A., Schöner, G., Riehle, A.: Preshaping and continuous evolution of motor cortical representations during movement preparation. The European Journal of Neuroscience **18**(7), 2047–2058 (2003)
3. Cisek, P., Kalaska, J.F.: Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. Neuron **45**(5), 801–814 (2005)
4. Conway, B.R., Tsao, D.Y.: Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. Proceedings of the National Academy of Sciences **106**(42), 18,034–18,039 (2009)
5. Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., Schöner, G.: The distribution of neuronal population activation (dpa) as a tool to study interaction and integration in cortical representations. Journal of neuroscience methods **94**(1), 53–66 (1999)
6. Fuster, J.M.: Cortex and mind: Unifying cognition. Oxford University Press, Oxford (2005)
7. Fuster, J.M., Alexander, G.E.: Neuron activity related to short-term memory. Science **173**(3997), 652–654 (1971)

8. van Gelder, T., Port, R.: It's about time: An overview of the dynamical approach to cognition. In: R. Port, T. van Gelder (eds.) Mind as motion, pp. 1–43. MIT Press, Cambridge and Massachusetts (1995)

9. Georgopoulos, A.P., Caminiti, R., Kalaska, J.F., Massey, J.T.: Spatial coding of movement: A hypothesis concerning the coding of movement direction by motor cortical populations. Experimental Brain Research **49**, 327–336 (1983)

10. Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., Massey, J.T.: On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. Journal of Neuroscience **2**(11), 1527–1537 (1982)

11. Georgopoulos, A.P., Kettner, R.E., Schwartz, A.B.: Primate motor cortex and free arm movements to visual targets in three-dimensional space. ii. coding of the direction of movement by a neuronal population. The Journal of Neuroscience **8**(8), 2928–2937 (1988)

12. Groh, J.M., Born, R.T., Newsome, W.T.: How is a sensory map read out? effects of microstimulation in visual area mt on saccades and smooth pursuit eye movements. The Journal of Neuroscience **17**(11), 4312–4330 (1997)

13. Grossberg, S.: A theory of human memory: self-organization and performance of sensory-motor codes, maps, and plans. In: F.M. Snell, R. Rosen (eds.) Progress in theoretical biology: Vol.5, pp. 500–639. Academic press, New York and London (1978)

14. Hock, H.S., Kogan, K., Espinoza, J.K.: Dynamic, state-dependent thresholds for the perception of single-element apparent motion: bistability from local cooperativity. Perception & psychophysics **59**(7), 1077–1088 (1997)

15. Hock, H.S., Schöner, G.: A neural basis for perceptual dynamics. In: R. Huys, V.K. Jirsa (eds.) Nonlinear dynamics in human behavior, pp. 151–177. Springer-Verlag, Berlin and Heidelberg (2010)

16. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. The Journal of Physiology **195**(1), 215–243 (1968)

17. Jancke, D., Erlhagen, W., Dinse, H.R., Akhavan, A.C., Giese, M., Steinhage, A., Schöner, G.: Parametric population representation of retinal location: neuronal interaction dynamics in cat primary visual cortex. The Journal of Neuroscience **19**(20), 9016–9028 (1999)

18. Johnson, J.S., Spencer, J.P., Luck, S.J., Schöner, G.: A dynamic neural field model of visual working memory and change detection. Psychological science **20**(5), 568–577 (2009)

19. Johnson, J.S., Spencer, J.P., Schöner, G.: A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. Brain research **1299**, 17–32 (2009). URL doi:10.1016/j.brainres.2009.07.008

20. Jones, J.P., Palmer, L.A.: The two-dimensional spatial structure of simple receptive fields in cat striate cortex. Journal of Neurophysiology **58**(6), 1187–1211 (1987)

21. Kopecz, K., Schöner, G.: Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. Biological cybernetics **73**(1), 49–60 (1995)

22. Lee, C., Rohrer, W.H., Sparks, D.L.: Population coding of saccadic eye movements by neurons in the superior colliculus. Nature **332**(6162), 357–360 (1988)

23. Luck, S.J., Vogel, E.K.: The capacity of visual working memory for features and conjunctions. Nature **390**(6657), 279–281 (1997)

24. Ottes, F.P., van Gisbergen, J.A., Eggermont, J.J.: Latency dependence of colour-based target vs nontarget discrimination by the saccadic system. Vision research **25**(6), 849–862 (1985)

25. Pasupathy, A., Connor, C.E.: Shape representation in area v4: position-specific tuning for boundary conformation. Journal of Neurophysiology **86**(5), 2505–2519 (2001)

26. Riegler, A.: When is a cognitive system embodied? Cognitive Systems Research **3**(3), 339–348 (2002)

27. Schöner, G.: Dynamical systems approaches to cognition. In: R. Sun (ed.) The Cambridge Handbook of Computational Psychology, pp. 101–126. Cambridge University Press, Cambridge (2008)

28. Schwartz, A.B., Kettner, R.E., Georgopoulos, A.P.: Primate motor cortex and free arm movements to visual targets in three-dimensional space. i. relations between single cell discharge and direction of movement. The Journal of Neuroscience **8**(8), 2913–2927 (1988)

29. Sherrington, C.S.: The integrative action of the nervous system. Yale University Press, New Haven (1906)
30. Simmering, V.R., Schutte, A.R., Spencer, J.P.: Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. Brain research **1202**, 68–86 (2008)
31. Spencer, J.P., Simmering, V.R., Schutte, A.R., Schöner, G.: What does theoretical neuroscience have to offer the study of behavioral development? insights from a dynamic field theory of spatial cognition. In: J.M. Plumert, J.P. Spencer (eds.) The emerging spatial mind, pp. 320–361. Oxford University Press, Oxford and New York (2007)
32. Trappenberg, T.P., Dorris, M.C., Munoz, D.P., Klein, R.M.: A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. Journal of cognitive neuroscience **13**(2), 256–271 (2001)
33. Wilimzig, C., Schneider, S., Schöner, G.: The time course of saccadic decision making: Dynamic field theory. Neural Networks **19**(8), 1059–1074 (2006)
34. Wilson, H.R., Cowan, J.D.: Excitatory and inhibitory interactions in localized populations of model neurons. Biophysical journal **12**(1), 1–24 (1972)
35. Zhang, W., Luck, S.J.: Discrete fixed-resolution representations in visual working memory. Nature **453**(7192), 233–235 (2008)